

Numerical evidence for nucleated self-assembly of DNA brick origami

Aleks Reinhardt and Daan Frenkel*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

(Dated: 26 February 2014)

The observation by Ke *et al.* [Science **338**, 1177 (2012)] that large numbers of short, pre-designed DNA strands can assemble into three-dimensional target structures came as a great surprise, as no colloidal self-assembling system has ever achieved the same degree of complexity. That failure seemed easy to rationalise: the larger the number of distinct building blocks, the higher the expected error rate for self-assembly. The experiments of Ke *et al.* have disproved this argument. Here, we report Monte Carlo simulations of the self-assembly of a DNA brick cube, comprising approximately 1000 distinct DNA strands, using a simple model. We model the DNA strands as lattice tetrahedra with attractive patches, the interaction strengths of which are computed using a standard thermodynamic model. We find that, within a narrow temperature window, the target structure assembles with high probability. Our simulations suggest that mis-assembly is disfavoured because of a slow nucleation step. As our model incorporates no aspect of DNA other than its binding properties, these simulations suggest that, with proper design of the building blocks, other systems, such as colloids, may also assemble into truly complex structures.

The development of DNA ‘origami’ [1–4] has made it possible to exploit the exquisite designability of DNA hybridisation to create a wide range of novel, self-assembling structures that promise to have applications in virtually all aspects of nanotechnology (for a recent review, see Ref. 5). The original version of DNA origami employed a long ‘scaffold’ single-stranded (ss)DNA sequence and linking ‘staple’ ssDNA molecules that serve to fold the scaffold strand into the desired shape [4]. A variety of structures have been assembled, including simple sheets, boxes that can open and close, ‘smiley faces’ and curved vase-like containers [6].

In 2012, Ke *et al.* reported a radically different approach that dispenses with the long ssDNA template [7]. The method of Ref. 7 is based on the pre-fabrication of small DNA bricks that can be linked together in a way somewhat akin to Lego bricks, but Lego bricks that fit in only one predetermined part of the target structure. With this approach, it proved possible to construct almost any target structure up to a given size simply by preparing a mixture of the designed DNA bricks and cooling it down. This modular approach makes the design of origamis considerably simpler than traditional DNA origami synthesis, in which a new set of staple strands must be designed for every new shape that one wishes to construct. Moreover, while traditional DNA origami takes a long scaffold strand from viral DNA, no biological DNA is required in DNA brick origami. Ke *et al.* demonstrated the applicability of their approach by constructing over 100 different shapes from a cuboid DNA ‘canvas’ [7], and this modular design has also been used to construct two-dimensional structures [8, 9] and more complex building blocks [10].

It should be stressed that the observation of Ref. 7 was very surprising. The self-assembly of short ssDNA strands may seem intuitive at first glance, given that DNA provides for precise sequence matching to allow only the correct ‘bricks’ to stick together, but in the self-assembly of (say) a molecular crystal, self-poisoning is a serious problem: if molecules are incorporated incorrectly in the crystal, the target structure cannot be reached. Apparently, DNA bricks manage to avoid this issue. This fact is even more surprising given that the bricks of Ref. 7 were made using ‘positive’ design only, whereby

the favourable interactions between putative neighbours were chosen, but no ‘negative’ design [3], in the sense that possible undesired interactions were not excluded. With many copies of each DNA strand in the system, the potential for incorrect assembly is significant. Indeed, templated DNA origami was developed precisely to avoid this problem [11]. Ke *et al.* suggest that in their system, seeding is slower than the subsequent growth of the desired structure, thereby minimising the tendency for incorrect assembly, but it is not immediately obvious that this should be the case [12].

The aim of the present Letter is to explore if a generic, and absolutely minimal, model of DNA bricks can reproduce the findings of Ke *et al.*: if this were to be the case, this would be good news, because it would imply that similar complex structures could be made with very different building blocks, as long as they had the same functionality as DNA bricks.

The basic principle of DNA origami design is that the target structure has the lowest free energy, something that is usually realised by maximising the number of complementary Watson–Crick base pairs (A–T or G–C) [13]. However, the kinetics of the self-assembly process are not well understood, and more specifically, we do not know what it takes to avoid kinetic traps [13]. Studying the DNA brick origami self-assembly process in detail would allow us to gain an understanding of the factors governing the rates and yields associated with the process of self-assembly and might eventually assist in the formulation of optimal design rules.

As DNA brick origami structures are extremely large, containing several thousand base pairs, all-atom simulations that would be long enough to observe self-assembly would be prohibitively time consuming. A coarse-grained model is therefore called for, but such a model, whilst simple, should not be *too* simple: it should capture the essential features of real DNA hybridisation. While several coarse-grained models have been developed in recent years [13], most of these are still much too detailed to be able to be used in a study of DNA brick origami assembly.

In order to decide on the principal physical features that need to be retained in a coarse-grained description of DNA suitable for assembling DNA brick origamis, we first consider some

aspects of the experimental system described in Ref. 7. In this system, each 32-nucleotide single-stranded DNA molecule bonds with four other DNA molecules through a quarter of its total length (called a ‘domain’) to form the final structure. Each double stranded segment thus comprises 8 base pairs, which gives a dihedral angle of approximately 90° [7]. Of particular interest is the property that, if we consider the centres of mass of each DNA single strand in the cubic DNA brick, these form a distorted diamond structure [7]. This suggests that we can describe each molecule, when bonded, as a tetrahedron to a first approximation. Therefore, in our approach, each single-stranded molecule is modelled as a particle with four distinct, tetrahedrally-arranged patches, and each of these patches has an associated DNA sequence.

We carry out our simulations on a cubic lattice with lattice parameter a . Particles interact if they are diagonally adjacent to each other, and the minimum distance between any two particles is $a\sqrt{3}$. Particle interactions are initially slightly repulsive ($\epsilon_{\text{init}}/k_B = 100$ K) to prevent large-scale agglomeration, but to this interaction energy we add the free energy of hybridisation for the longest complementary ($5'-3'/3'-5'$) sequence match between the closest pair of ‘patches’, allowing for single internal mismatches. This free energy is determined using the nearest-neighbour parameterisation of SantaLucia Jr and co-workers [14], where we take into account the nearest neighbour sequences, terminal A-T penalties, internal mismatches [15] and dangling ends [16], as well as the temperature and salt concentration dependence [17]. However, we do not consider loops or bulges, which we do not expect to be important for sequences of at most 8 base pairs. We perform Metropolis Monte Carlo simulations [18] in the canonical ensemble with a periodic simulation box. To allow for more efficient sampling, we allow clusters of particles to be moved or rotated concurrently by using the virtual move Monte Carlo algorithm [19]. Particles (or clusters) are randomly translated along the lattice or rotated in one of a fixed number of predetermined rotations which account for the 24 possible orientations associated with a tetrahedron with 4 distinct vertices placed within a cube. To find the free energy as a function of the size of the largest correctly-bonded cluster, we run umbrella sampling [20] simulations with umbrella sampling steps performed every 200 000 Monte Carlo steps [21].

Our target structure comprises 998 single-stranded DNA molecules that, when correctly assembled, should form a cube (Fig. 1). Ke *et al.* found in their experimental work that randomly selected sequences that fulfil the bonding requirements result in yields comparable to those obtained by using specially optimised DNA sequences [7], and in the light of this, we have selected a random set of sequences for the patches in the target structure, but such that patches that are adjacent (*i.e.* bonded) in the correctly-assembled structure have complementary sequences. Like the ‘protector bricks’ in experiment [7], any patches at the boundaries of the cube that do not interact with another particle are given a sequence of 8 consecutive thymines to minimise the chance of their misbonding with other parts of the structure. Of the 998 particles simulated, 24 have only a single interaction with the remaining structure, and are unlikely

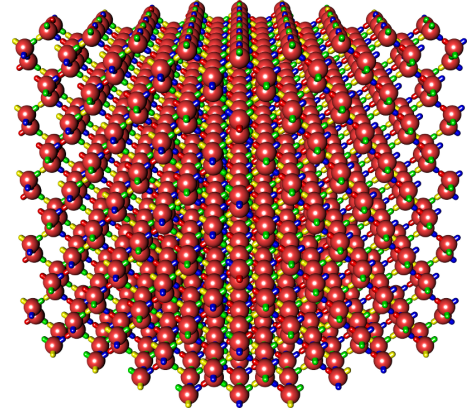


FIG. 1. (Colour online) Target structure. Each patch is colour-coded: by design, red patches bond with blue ones and green patches with yellow ones, but each patch has its own sequence.

to be able to form stable bonds.

Having determined the appropriate patch sequences, we run Monte Carlo simulations starting from a gas of monomers corresponding to a single target structure across a range of temperatures [22]. At high temperatures, any clusters that form are transient and small. At temperatures around 320 K, however, we observe very interesting behaviour. Several configurations along a particular trajectory at 318 K are shown in Fig. 2. It is clear from this figure that the system is able to assemble into the designed structure at this temperature. Moreover, we see that several other clusters (which we define as comprising particles connected by at least one bond corresponding to the designed structure) do grow in addition to the largest one, sometimes connected to the largest one and sometimes not, but at this temperature, they are not sufficiently stable to persist and only one cluster grows at the expense of all others. If we decrease the simulation temperature (Fig. 3), we find ever larger aggregates of different, incorrectly-bonded clusters, *i.e.* clusters in which patch bonding is not perfectly complementary. The size of the largest cluster in the system for a particular set of simulations at different temperatures as a function of Monte Carlo time is shown in Fig. 4; we can see that at high temperatures, no clusters form; at intermediate temperatures, clusters can grow to large sizes; and at low temperatures, the largest cluster does not grow considerably after an initial growth stage, as other clusters have formed and misbonded, and these ‘incorrect’ bonds do not dissociate readily. At temperatures just under the successful assembly regime, the largest cluster can grow to appreciable sizes, but more than one large correctly-bonded cluster typically forms and these clusters then struggle to meet in the correct way, resulting in a misformed structure. If we run simulations starting from the fully formed structure, the structure remains mainly intact to temperatures between 325 K and 330 K; in other words, there is some degree of hysteresis in the transformation.

We have simulated three additional independent repeats of the simulations discussed above at equidistant temperatures between 310 K and 325 K inclusive, and a further 10 runs each for temperatures between 317 K and 319 K, and observed the

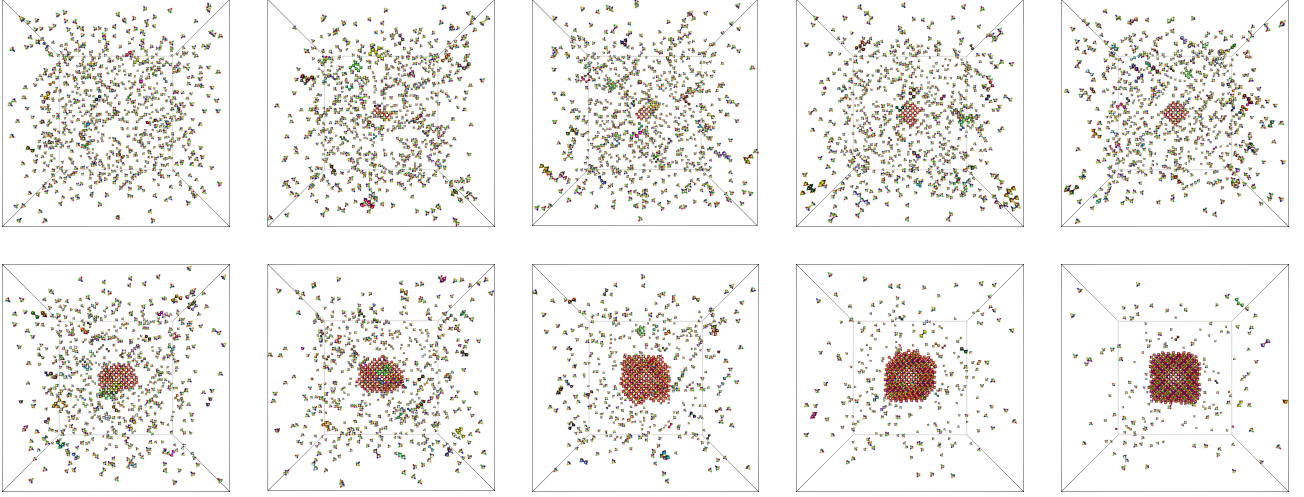


FIG. 2. (Colour online) Simulation snapshots from a single trajectory at $T = 318$ K. Snapshots were taken approximately every 2×10^{10} Monte Carlo steps and are arranged in sequence from left to right and from top to bottom in the figure. The largest correctly-bonded cluster is shown in red at the centre of the simulation box; other large correctly-bonded clusters are shown in other colours.

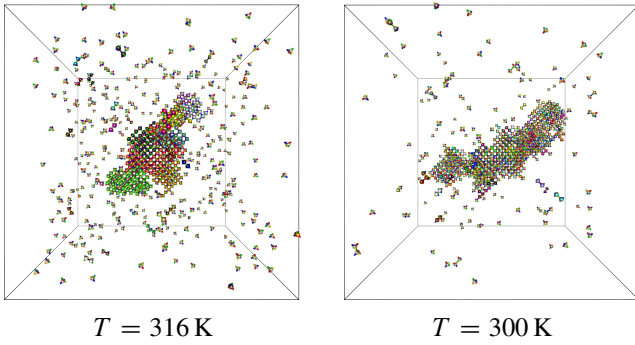


FIG. 3. (Colour online) Two simulation snapshots from lower-temperature simulations show the formation of kinetic aggregates.

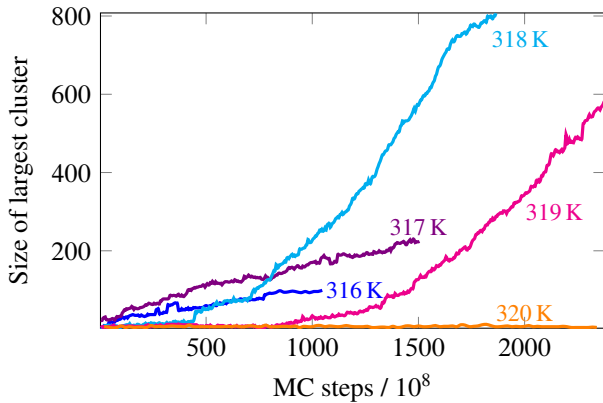


FIG. 4. (Colour online) The size of the largest correctly-bonded cluster is shown as a function of the number of Monte Carlo steps for a range of temperatures for a particular set of trajectories.

same qualitative behaviour in all additional runs. The correct structure forms at temperatures between about 317 K and 319 K, but with various lead times before significant growth takes place. This suggests that there is a free-energy barrier to nucleation that increases with the temperature; the higher the

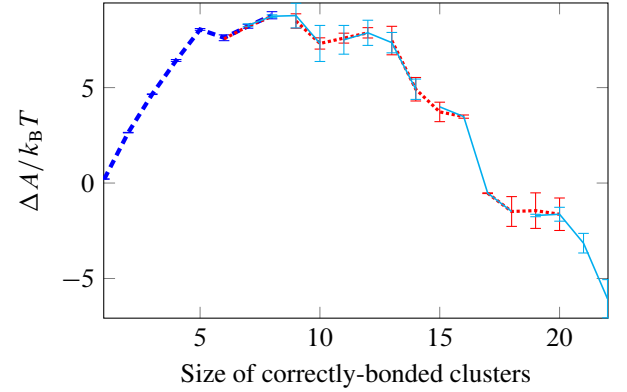


FIG. 5. (Colour online) The free-energy profile for cluster growth as a function of the size of correctly-bonded clusters at $T = 319.5$ K. Simulation results from different umbrella sampling windows are depicted in alternating shadings and line styles to show their overlap. The blue dashed line corresponds to brute-force simulations.

temperature, the rarer the nucleation event, but, by contrast, the smaller the chance of incorrect assembly. In order to quantify the magnitude of this free-energy barrier, we have run umbrella sampling simulations at 319.5 K, where the free-energy barrier is expected to be relatively small. We plot the free energy as a function of the cluster size in Fig. 5 [23]. The number of clusters of a given size per unit volume decreases rapidly with the cluster size, and in order for the largest cluster in the system to grow beyond just a few particles, a free-energy barrier must be overcome. The critical cluster size at 319.5 K is approximately 10; beyond this size, the free energy predominantly decreases as the cluster grows. However, this decrease is not monotonic, reflecting the fact that certain clusters, which tend to involve ‘caged’ structures with few dangling particles, are favoured over others; this is not dissimilar to the multi-peaked nucleation barriers seen in Ising-type models [24]. Nevertheless, the principal free-energy barrier to nucleation at this temperature appears to be relatively small, consistent with the fact that

spontaneous growth is (eventually) observed in brute-force simulations. Moreover, brute-force simulations starting from the critical cluster size as determined by umbrella sampling simulations confirm that the critical cluster size has been correctly identified, although the precise structure of a given cluster has a significant effect when considering its propensity to grow or shrink: the cluster size alone is not an optimal order parameter.

We have also performed several additional brute-force simulations with a different random choice of patch sequences; the same overall behaviour is observed, although the precise temperature range at which nucleation occurs varies by a few degrees. Nevertheless, it appears that regardless of the choice of sequence, a slow annealing process from high temperatures will result in the growth of the designed structure, as the system will always pass through the optimal growth regime on cooling.

It is intriguing that the designed structures nucleate reproducibly; however, it is worth looking at the limits of the model and the effects that we have neglected. Firstly, some of the most competitive alternative structures are likely to be ones that form with exactly the correct sequence pairing, but with different replicas of the same molecules being incorporated into the same final structure, which would disrupt the geometry of the growing cluster and lead to a frustrated system. On a lattice, this becomes much less probable because the geometry in which the system grows is essentially externally imposed; furthermore, we have, at this stage, not attempted to grow multiple copies of the final structure concurrently. Furthermore, even in the absence of this competition effect, simulating the growth of a single target structure is unlikely to result in bulk assembly statistics [25]. Secondly, the fact that we model the system as a ‘patchy’ potential has several implications. Single-stranded DNA particles have a reduced entropy relative to the experimental system, because we fix the tetrahedral geometry in advance. This will in all probability lead to a relative destabilisation of the single-stranded system and will mean that any melting points we may obtain are expected to be higher than in experiment. Moreover, we keep the bond angles fixed irrespective of the number of complementary bases between two patches; that is, the bond angle does not change when we have an interaction involving fewer than 8 base pairs. We further do not consider any hybridisation between parts of domains (for example, in reality, a strand might bond preferentially with part of domain 1 and part of domain 2 of another strand, but we only consider bonding with either domain 1 or domain 2). We also do not correct for the fact that, within our model, there may be several bonding patterns with fewer than 8 matching base pairs of similar strength possible between a pair of patches, which might be expected to stabilise some weak bonds entropically. Finally, although the 48-nucleotide-long ‘boundary bricks’ seem to be important in experiment, and are likely to be more important for more intricate structures than cubes, we have not considered them explicitly in our simulations. However, whilst it is certainly important to be aware of these simplifications and omissions in our simulations, the basic underlying physical behaviour of the self-assembly process appears to be captured by our model, and our simulations

support the suggestion of Ke *et al.* that initial structure growth is a slow process.

In summary, we have performed lattice Monte Carlo simulations of a model system designed to mimic the behaviour of DNA bricks studied experimentally by Ke *et al.* [7]. We have demonstrated that there is a sweet spot in temperature for which the self-assembly of the target structure is successful. Above this temperature range, the monomer phase is entropically favoured, while below it, non-specific bonding results in the growth of large aggregate structures. In experiment, structures are formed via a slow annealing process, passing through the optimum temperature range, and so the desired structures form in reasonable yields. There appears to be a free-energy barrier to nucleation at higher temperatures, and so our simulations support the basic premise that slow nucleation is followed by faster structure growth, as posited by Ke *et al.* [7]. Finally, the very fact that we use a highly simplified model implies that our results should carry over to other systems, such as (nano)colloids that have been designed with the same properties as DNA bricks. This observation is extremely encouraging, because it suggests that it should be possible to assemble structures consisting of materials other than DNA into complex target structures.

This work was supported by the European Research Council [Advanced Grant 227758] and the Engineering and Physical Sciences Research Council [Programme Grant EP/I001352/1]. We thank Thomas Ouldridge for useful discussions.

* Correspondence author. E-mail: df246@cam.ac.uk

- [1] N. C. Seeman, *Nature* **421**, 427 (2003).
- [2] N. C. Seeman, *J. Theor. Biol.* **99**, 237 (1982).
- [3] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, *Nature* **394**, 539 (1998).
- [4] P. W. K. Rothmund, *Nature* **440**, 297 (2006).
- [5] V. Linko and H. Dietz, *Curr. Opin. Biotech.* **24**, 555 (2013).
- [6] T. Tørring, N. V. Voigt, J. Nangreave, H. Yan, and K. V. Gothelf, *Chem. Soc. Rev.* **40**, 5636 (2011).
- [7] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, *Science* **338**, 1177 (2012).
- [8] B. Wei, M. Dai, and P. Yin, *Nature* **485**, 623 (2012).
- [9] Z. Zhang, J. Song, F. Besenbacher, M. Dong, and K. V. Gothelf, *Angew. Chem., Int. Ed.* **52**, 9219 (2013).
- [10] B. Wei, M. Dai, C. Myhrvold, Y. Ke, R. Jungmann, and P. Yin, *J. Am. Chem. Soc.* **135**, 18080 (2013).
- [11] P. W. K. Rothmund and E. S. Andersen, *Nature* **485**, 584 (2012).
- [12] K. V. Gothelf, *Science* **338**, 1159 (2012).
- [13] J. P. K. Doye, T. E. Ouldridge, A. A. Louis, F. Romano, P. Šulc, C. Matek, B. E. K. Snodin, L. Rovigatti, J. S. Schreck, R. M. Harrison, and W. P. J. Smith, *Phys. Chem. Chem. Phys.* **15**, 20395 (2013).
- [14] J. SantaLucia Jr and D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415 (2004).
- [15] H. T. Allawi and J. SantaLucia Jr, *Biochemistry* **36**, 10581 (1997); *Nucl. Acids Res.* **26**, 2694 (1998); *Biochemistry* **37**, 9435 (1998); *Biochemistry* **37**, 2170 (1998); N. Peyret, P. A. Seneviratne, H. T. Allawi, and J. SantaLucia Jr, *Biochemistry* **38**, 3468 (1999).
- [16] S. Bommarito, N. Peyret, and J. SantaLucia Jr, *Nucl. Acids Res.*

- 28**, 1929 (2000).
- [17] R. T. Koehler and N. Peyret, *Bioinformatics* **21**, 3333 (2005).
- [18] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [19] S. Whitelam and P. L. Geissler, *J. Chem. Phys.* **127**, 154101 (2007); *J. Chem. Phys.* **128**, 219901 (2008).
- [20] G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977); M. Mezei, *J. Comput. Phys.* **68**, 237 (1987).
- [21] B. Hetényi, K. Bernacki, and B. J. Berne, *J. Chem. Phys.* **117**, 8203 (2002).
- [22] The salt concentrations used for the nearest-neighbour sequence-dependent potential were $[\text{Na}^+] = 1 \text{ mol dm}^{-3}$ and $[\text{Mg}^{2+}] = 0.08 \text{ mol dm}^{-3}$. These values are slightly different from the experimental setup of Ke *et al.*, but are in the range where the salt concentration dependence formula given by Koehler and Peyret [17] is applicable. In the simulations reported here, the lattice dimensions were $(62a)^3$ and the simulation box was periodic; the simulation box volume is of course important when mapping the results to experiment and can significantly affect the nucleation rate, but as we are not directly comparing to a specific experiment, this is an ‘arbitrary’ parameter in our simulations at this stage.
- [23] We have defined the free energy as $\Delta A(n)/k_B T \equiv -\ln(N_n/N) \propto -\ln(N_n/V)$, where N_n is the number of clusters of size n and N is the number of particles in the system. For small n , this can be estimated directly from brute-force simulations of the initial state of the system. For larger n , N_n/N becomes small and, moreover, rapidly decreases with n . It therefore approaches the probability that the largest cluster in the system is of size n . Beyond $n = 7$, we assume that these probabilities are the same, and we calculate free energies using umbrella sampling with the size of the largest cluster serving as the order parameter.
- [24] V. A. Shneidman, *J. Stat. Phys.* **112**, 293 (2003).
- [25] T. E. Ouldridge, *J. Chem. Phys.* **137**, 144105 (2012).